

## MSc/PGDip in Data Analytics (January 2023 Entry)

### Modelling, Simulation and Optimisation (H9MSO)

#### Project CA

**DEADLINE:** 7<sup>th</sup> August 2023, 23:59

**WEIGHT:** 60% of overall marks

#### Project Outline

‘We-Doo’ is a start-up company whose business idea is to take over the delivery services in commuter townships that grew up all over the country. Essentially they allow the normal delivery services to drop their packages in a secure location in the township and in the evening a local driver would distribute these packages on an electric cargo bike to the end customers.

To validate their business model ‘We-Doo’ has task you to simulate the typical evening delivery runs for a small township.

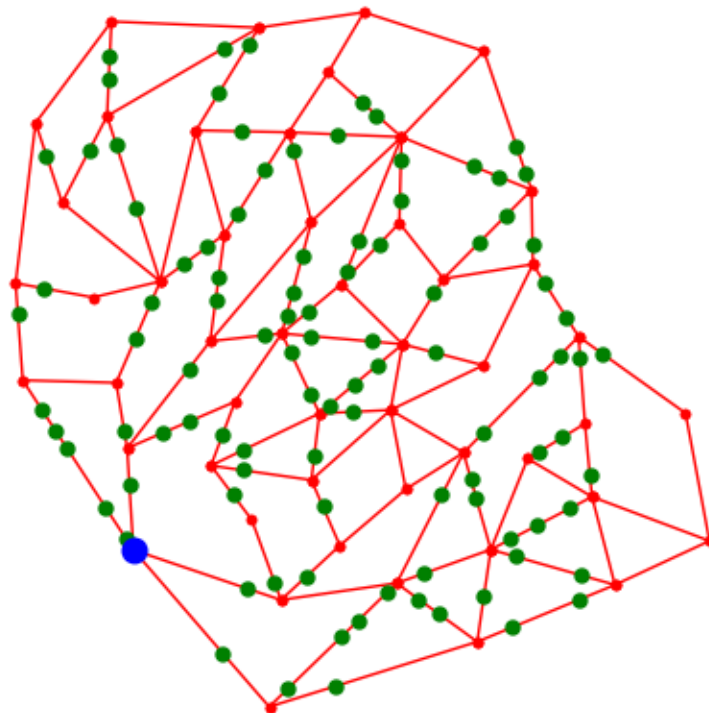
#### Parameters

You are given the following information:

- A map  $M$  of the township with the location of the local distribution point  $W$  and a list of customer locations  $C$ . Figure 1 shows an example with 100 customer locations.
- The average number  $p$  of parcels a customer receives per day is not exactly known. It is assumed to be equal for all customers and constant over time. It is assumed that the value is within the range  $0.2 \leq p \leq 0.3$ . The interarrival time for parcels for a given customer is an independent expovariate distribution with  $\mu = 1/p$ . Figure 2 shows a sample distribution for the interarrival time. Figure 3 shows the cumulative number of parcels received per day and per customer for 100 customers over 50 days, both for  $p = 0.2$ .
- The company has a choice between electric cargo bikes with a range of 30, 35, 40 or 45km. In the long run it is probably the best to use cargo bikes with 45km range because they give more flexibility and they are also supposed to last longer. At the moment the company is cash strapped. The company wants to test the business idea with the cheapest possible bike and the price of the bike grows substantially with the range.
- The delivery driver has a planned average working time of 3 hours per working day (18:00-21:00).

One of the founders has worked himself in a previous life as a delivery driver and provides the following additional information:

- The time required for driving can be estimated based on the path length between any two points at an average speed of 15km/h.
- The time for handing over the parcels to the customer consists of a call time (for the customer to come to the door) that follows an expovariate distribution with a mean time of 40sec) and an additional handover time per parcel, which again follows an exponential distribution with on average 10s per parcel.
- The cumulative preparation time (route planning and sorting of the parcels in the delivery order and packing the cargo-bike) is assumed to be 50 sec per parcel to be delivered.
- The day-end procedure (reporting and setting up the cargo-bike for charging) is 10 minutes.



*Figure 1 Sample street map with distribution point (blue) and 100 customer locations (green)*

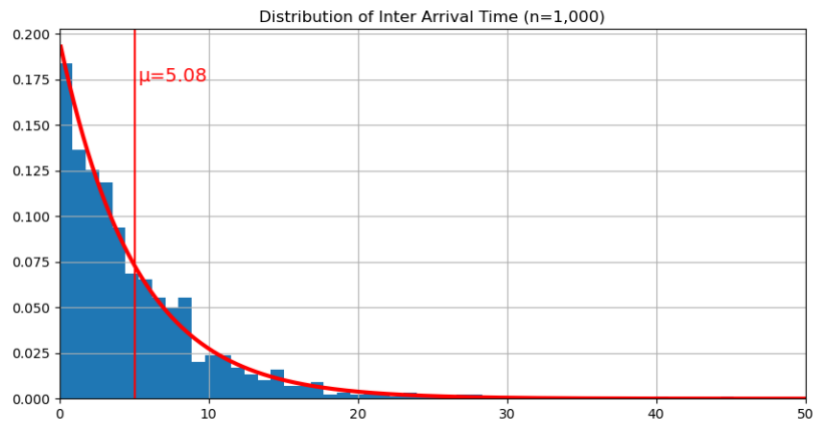


Figure 1

Figure 2 Interarrival Time of parcels for a single customer with  $p=0.3$ .



Figure 3 Resulting sample distribution of the number of parcels received per customer and per day for  $p=0.2$ .

## Input Data Provided

You are provided with a pickled file **data.pickled** containing M, W, and C. This can be loaded with

```
import pickle
with open('data.pickled', 'rb') as f:
    M, W, C = pickle.load(f)
```

For model verification and early test runs you may want to generate smaller townships or customer numbers. For this purpose you are provided with a Jupyter Notebook file **GenerateData.ipynb** which contains a method **generateData** which can be used to generate other sample data. Call **generateData()** without parameters and it will print a help text explaining the parameters and output values.

## The Problem

Your task is to provide a simulation solution that can in principle be used to plan the rollout for new townships, but in the moment the company needs advice which of the preselected cargo bikes they should use for the trial.

- As the average number  $p$  of parcels a customer receives per day is not currently known, you need to report for a number of values in the assumed range  $0.2 \leq p \leq 0.3$  but possibly also for values outside this range. It is your decision how many and which values you will investigate. Choose at least 3 values.
- For each of the  $p$  values you have chosen, you need to investigate the performance for the cargo bike with different ranges  $r$  (30km, 35km, 40km, and 45km) and argue for your recommendation for the choice of cargo bike.

## Reporting of Results

In your final report you need to report for each of the values  $p$  and  $r$  on the following :

1. Report on the **daily tour length**, as shown below in figure 4 and 5.  
When the actual tour length reaches the bikes range it shows that there is no room of delivering more parcels. It may just be enough, but it could result in parcels being left over for the next day and hence a delivery delay. Keep in mind that this is a *hard constraint*, it should not be broken on any day.

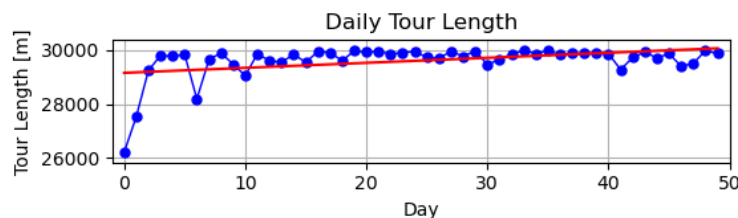


Figure 4. Plot of daily tour length over 50 working days (30 km bike range)

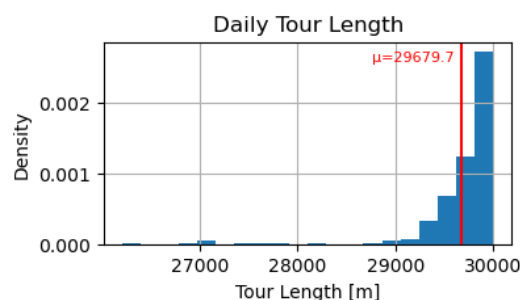


Figure 5. Histogram of daily tour length (30km bike range)

2. Report on the **daily working time**, as shown below in figure 6 and 7.  
When using a cargo bike with 40km range, the constraints on the delivery capacity relax, but at the same the working time is getting longer. The working time is a *soft*

*constraint.* A certain overrun on some days could be acceptable, provided the average works out.

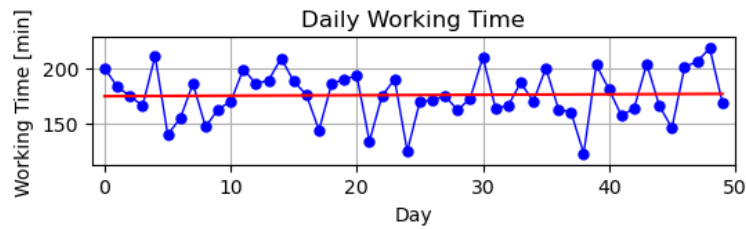


Figure 6. Plot of daily working time over 50 days (40 km bike range)

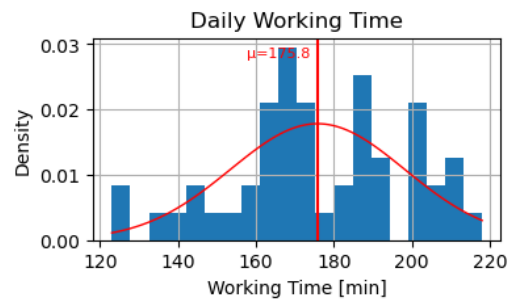


Figure 7. Histogram of daily working time (40 km bike range)

- Report on a daily basis on the **number of left over parcels**, as shown in figure 8 and 9. Should the daily plot show a statistically significant trend with an increasing number of left over parcels, the model is clearly not sustainable. A bit of a daily left over may on the other side smooth out variation in actual daily arrival numbers.

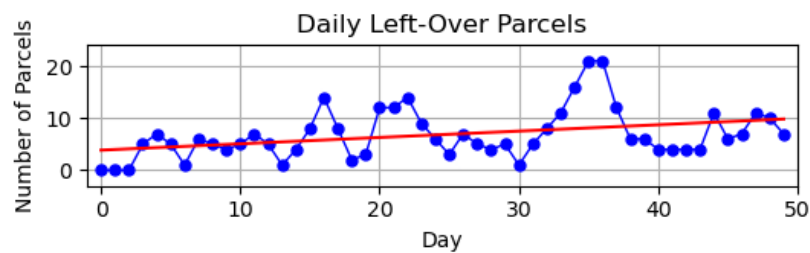


Figure 8. Plot of Daily Left Overs per Day (30km bike range)



Figure 9. Histogram of daily left over parcels (30km bike range)

4. Report on the delay time per parcel, as shown in table 1.  
This is a critical indicator of the quality of the service that can be provided, which will influence the final business decision.

Delivery Delay	p=0.2 30km bike range	p=0.2 40km bike range
None	61.2%	98.9%
1 day	37.5%	1.1%
2 days	1.2%	0%
> 2 days	0%	0%

*Table 1. Sample Delivery Delay per Parcel depending on bike range (p=0.2)*

## Deliverables

The project has two deliverables:

1. A final report in .pdf format describing your simulation study should have maximal 6 pages in IEEE conference format and follow the outline structure given below
2. A .zip file that contains the code as (one or more) Jupyter Notebook files and any additional data, configuration files or documentation that may be required to run the code.

The final report will be uploaded through the Turn-it-in link provided on Moodle.

For evaluation criteria please check the Rubrics given at the end of this project outline.

You may use components from the Simulation Study presented in class. When re-using code or design elements, refer to the version uploaded on Moodle.

## Academic Integrity

- By submitting your work on Moodle you declare that this is your own work.
- Any material created by others must be properly referenced. Verbatim text copies should be included in quotes.
- All figures included in the report should have their origin in one of the jupyter notebooks uploaded. Provide the file name and cell number as part of the figure caption. For figures not created by yourself the caption should include an acknowledgement detailing the name(s) of the creator(s) and proper references.
- Code copied from class material or other sources should be clearly marked as such and properly referenced. In particular it should not be (directly or implicitly) claimed as your own. Instead a comment should be included in the source code indicating where you obtained it from. If your modification forms just a small part of the code, highlight explicitly what is yours.
- Students are strongly advised to familiarise themselves with the Guide to Academic Integrity. All submissions will be electronically screened for evidence of academic misconduct, e.g. plagiarism, collusion and misrepresentation. Any submission showing evidence of such misconduct will be referred to the college's processes.

## The Structure of the Final Report

### Title

#### Abstract with Keywords

##### 1. Introduction

In this section you introduce the problem you intend to investigate and document the parameters you are using for your study as they have been generated by the **generateData()** method from the last four digits of your student ID.

##### 2. Literature Review

To the extent you have referred to the literature for your simulation study give a summary of your sources and the ideas you have extracted. When you refer to data provided in class use as reference the material provided on Moodle.

##### 3. Methodology

Describe the sequence of your development and give references to the relevant section of code in the Jupyter Notebook files. Give details of the model verification and validation you performed and the number of simulation runs, the load generator function used, and the statistical evaluation of the simulation results.

##### 4. Results and Interpretation

Report the results of your simulation study for the planned daily delivery capacity and the maximal sustainable daily delivery capacity and give an interpretation of the same.

##### 5. Reflections and Future Work

Discuss how your research could be improved and suggest problems for future research.

##### 6. References

## Rubrics

Grade Criterion	Solid H1 > 80%	H1 > 70%	H2.1 > 60%	H2.2 > 50%	PASS > 40%	FAIL < 40%
<b>Methodology (20%)</b>	All elements of project requirements have been thoroughly addressed. The logic of the simulation study is well presented.	All elements of the project requirements have been thoroughly addressed. Arguments have been given for the type and number of simulation runs.	Some minor requirements missing from project. No arguments have been given for the type and number of simulation runs.	Multiple omissions from the project. The proposed number of simulation runs is insufficient for statistical evaluation.	Major parts of the project are missing. The project may contain parts that are not relevant. The proposed simulation runs are insufficient for the project.	The solution bears no resemblance to the project requirements at all.
<b>Simulation (20%)</b>	An excellent, thorough simulation was carried out. Effort exceeds the requirements.	An excellent, fully complete simulation was carried out. The results go beyond the minimal requirement.	A very good and largely complete simulation of the required models was carried out.	A good and largely complete simulation of one model or an inadequate simulation of multiple models was carried out.	An inadequate simulation of one model was carried out. Some logical errors exist.	Little or no simulation carried out.
<b>Code (20%)</b>	The notebook executes without problems. Code is elegant and fully commented. The notebook is well presented. The implementation significantly exceeds the requirements	The notebook executes without problems. Code is fully commented. There is no excess code used. The notebook is well presented.	The notebook executes without problems. Code is partially commented. There is a minimal amount of excess code used.	The notebook works and allows the output to be reproduced. The notebook is poorly commented or contains a lot of excess code.	The output of the notebook file can't be reproduced. The notebook file contains minor errors, is not commented or contains a lot of unrelated code.	The notebook file doesn't contain output or contains errors that prevent it from being executed.
<b>Evaluation &amp; Results (30%)</b>	Models are fully evaluated. Results are thoroughly discussed. Statistics is applied flawlessly. There is significant reflection on the challenges faced in this project and possible resolution to remaining problems	Models are fully evaluated. Results are presented and thoroughly discussed. Statistics is applied appropriately. There is significant reflection on the challenges faced in this project.	Models are evaluated based on a sufficient number of simulation runs. Results are presented and thoroughly discussed. There is very good reflection on the challenges faced in this project.	Models and results are presented and appropriately discussed. There is good reflection on the challenges faced in this project in particular related to the insufficient number of simulation runs, if this was the case.	Cursory evaluation of one model. Cursory discussion of the results. There is some reflection on the challenges faced in this project. There is no reflection on the insufficient number of simulation runs.	Little to no evaluation of model. Little to no discussion of results. There is no reflection on the challenges faced in this project
<b>Quality of Writing (10%)</b>	Very well written, with no language errors. All figures are well conceived and readable. The IEEE template is strictly adhered to. Report does not exceed the length limits. References are appropriately and correctly used.	Well written, with only minor language errors. All figures are well conceived and readable. The IEEE template is adhered to. Report does not exceed the length limits. References are appropriately and correctly used.	Main document has a few language and/or style errors. Figures are well presented. IEEE template and length limit are adhered to. References are complete, and correctly used.	Main document has a few language and/or style errors. Some figures are may be hard to read. IEEE template and length limit are largely adhered to. References are complete, and correctly used.	Main report is readable with some language and/or style errors. Figures may be hard to read or presented in a suboptimal manner. IEEE template may have been broken. References are mostly complete and correctly used.	Littered with typos, and/or poor use of English. IEEE template not used. Figures may be hard to read. References (if any) are incomplete or incorrect.